Revista Brasileira de Ciência do Solo

Division - Soil Use and Management | Commission - Soil and Water Management and Conservation

Sediment source tracing in a Brazilian subtropical catchment using diffuse reflectance: Effect of spectral ranges, pre-processing techniques, and multivariate model

Rafael Ramon[1,2], Olivier Evrard[3], Jean Paolo Gomes Minella[4], Cláudia Alessandra Peixoto de Barros[5], Jean Michel Moura-Bueno[4], Gabriela Naibo[5], Laurent Caner[6], Danilo Santos Rheinheimer[4] and Tales Tiecher[5]*

[1] Universidade Federal do Rio Grande do Sul, Departamento de Solos, Programa de Pós-Graduação em Ciência do Solo, Porto Alegre, Rio Grande do Sul, Brasil.
[2] BASF SA, Environmental Fate - Regulatory Science Crop Protection, São Paulo, São Paulo, Brasil.
[3] Université Paris-Saclay, Laboratoire des Sciences et de l'Environnement, Gif-sur-Yvette Cedex, Ilê-de-France, France.
[4] Universidade Federal de Santa Maria, Departamento de Ciência do Solo, Santa Maria, Rio Grande do Sul, Brasil.
[5] Universidade Federal do Rio Grande do Sul, Departamento de Ciência do Solo, Porto Alegre, Rio Grande do Sul, Brasil.
[6] Université de Poitiers, Poitiers, Nouvelle Aquitaine, France.

*** Corresponding author:**
E-mail: tales.tiecher@ufrgs.br

Received: November 29, 2023

Approved: March 15, 2024

How to cite: Ramon R, Evrard O, Minella JPG, Barros CAP, Moura-Bueno JM, Naibo G, Caner L, Rheinheimer DS, Tiecher T. Sediment source tracing in a Brazilian subtropical catchment using diffuse reflectance: Effect of spectral ranges, pre-processing techniques, and multivariate model. Rev Bras Cienc Solo. 2024;48:e0230144.
https://doi.org/10.36783/18069657rbcs20230144

Editors: José Miguel Reichert and Jeferson Dieckow.

ABSTRACT: Agriculture intensification in Southern Brazil's subtropical regions combined with the frequent occurrence of erosive rainfall has rendered the area a global water erosion hotspot. In this scenario, understanding and regulating erosion processes at the river catchment scale is critical for mitigating soil and water resource degradation. Traditional methods for tracing sediment sources are expensive and time-consuming and justify the development of alternative approaches. Therefore, in this study, we employed diffuse reflectance spectroscopy analyses in the ultraviolet-visible (UV-VIS), near-infrared (NIR), and mid-infrared (MIR) ranges, combined with multivariate models and spectral pre-processing techniques to estimate sediment source contributions in a homogeneous subtropical catchment (Conceição River, 804 km^2). Soil samples (n = 181) were collected to characterize the four potential sediment sources, including: cropland (n = 78), stream bank (n = 36), unpaved road (n = 40) and pasture (n = 27). Moreover, 44 sediment samples were collected, including suspended sediment (n = 8), fine sediment deposited on the riverbed (n = 15), and suspended sediment samples collected in the water column during storm events (n = 21). Vector machine (SVM) model outperformed the others, with better accuracy and reliability. While UV-VIS spectra proved less effective due to soil homogeneity across the catchment, NIR and MIR spectra provided valuable information for discriminating sediment sources. Furthermore, reducing the number of potential sources (from four to three or two) improved model predictions, especially when distinguishing between surface sources (cropland and pasture) and subsurface sources (unpaved roads and stream banks). The study's findings shed light on the power of efficient and cost-effective alternative methods for assessing sediment sources, which are vital for promoting effective erosion control and sustainable land management in similar regions.

Keywords: erosion processes, spectral discrimination, sediment fingerprinting, support vector machine, land management.

INTRODUCTION

Agriculture intensification without effective soil conservation management practices (often referred to locally as 'conservationist') associated with the high rainfall erosivity observed in subtropical regions of Southern Brazil make the region a global water erosion hotspot (Golosov and Walling, 2019). Therefore, measures to control erosion processes need to be adequately designed to mitigate soil and water resource degradation. At the river catchment scale, identifying and quantifying the contribution of the main sediment sources delivering material to the river system helps guide the efforts to control erosion processes more effectively (Collins et al., 2017).

Although monitoring studies of the impacts of human activity and climate change on natural resources (soil and water) is essential (Poesen, 2017), this type of research has remained relatively scarce in Brazil (Melo et al., 2020). Conventional techniques for tracing the sources of sediments delivered to the river network using tracers such as radionuclides, stable isotopes of specific organic compounds, or even geochemical composition are expensive and time-consuming methods (Collins et al., 2020).

In this context, diffuse reflectance spectroscopy in the visible and infrared wavelengths has emerged as a potential alternative to conventional analytical methods. This method relies on physical measurements that do not generate chemical residues since it requires only a small sample quantity for analysis and generates a large amount of information about the sample composition (Viscarra Rossel et al., 2006). Spectroscopy data has already been used to derive parameters related to soil organic constituents and mineral composition (Amorim et al., 2021), as well as color parameters (Tiecher et al., 2015; Pulley et al., 2018; Sellier et al., 2021) to be used as discrete variable tracers in mixture models. Another option is to use all spectra as a set of potential fingerprint properties through multivariate models. This method has already been used successfully to predict the contribution of sediment sources in contrasted catchments of Europe (Poulenard et al., 2009, 2012). In small heterogeneous catchments, the potential of this technique has also been demonstrated in Southern Brazil (Tiecher et al., 2016, 2021). In these studies, the magnitude of the source contributions to sediment predicted by the alternative method based on spectroscopy data was similar to those obtained by the conventional method based on geochemical tracers.

In contrast, the method has not yet been tested in a context where soil classes are homogeneous and more than two or three potential sediment sources may be found, such as in the Conceição river catchment in Southern Brazil (Ramon et al., 2020). Several studies have been conducted using diffuse reflectance spectroscopy as tracers and multivariate models to predict source contributions, as well outlined in the literature review of 28 scientific papers on this topic presented by Tiecher et al. (2021). However, they were conducted in smaller catchments with more contrasting characteristics of the sources. In this context, the combination of multiple spectral bands such as ultraviolet-visible (UV-VIS), near-infrared (NIR) and mid-infrared (MIR) may provide an alternative to improve the accuracy of model estimates. The use of multiple spectral ranges may improve the model performance due to the different types of information regarding the soil and sediment properties contained in each region of the spectra (Viscarra Rossel et al., 2006). The UV-VIS spectra provide mainly information related to the organic matter and iron oxide content, while the MIR and NIR wavelengths may provide information related to the occurrence of different minerals types and organic matter compounds, which together may result in a wide set of contrasted properties (Tiecher et al., 2021; Viscarra Rossel and Behrens, 2010). In addition, spectral pre-processing techniques for parametric and non-parametric multivariate models, such as Partial Least Squares Regression (PLSR) and Support Vector Machine (SVM), may improve the accuracy of results by removing baseline shifts and enhancing spectral features. This strategy had already been applied successfully by Tiecher et al. (2021) in a small catchment

(1.23 km²) of Southern Brazil, where the SVM and PLSR associated with the first derivative of Savitzky-Golay spectral pre-processing offered the best results. However, this approach remains to be tested and validated in larger and more homogeneous catchments, especially when considering multiple potential sediment sources.

This research aimed to test and compare the performance of PLSR and SVM models for estimating the contribution of sediment sources to the river network in a homogeneous agricultural catchment using diffuse reflectance spectroscopy outputs as a set of potential fingerprints. In addition, our goal is to verify which spectral range and which pre-processing technique offers the best results for sediment source contribution predictions in a large catchment with highly weathered and homogeneous soils in Southern Brazil, as this type of environment is considered nowadays as one of the soil erosion hotspots of the world (Foucher et al., 2023).

MATERIALS AND METHODS

Study site

The study was conducted in the Conceição River catchment (804 km²), located in the northwest of the state of Rio Grande do Sul (Figure 1). According to the Köppen classification system, the climate is classified as Cfa, humid subtropical without a defined dry season, with an average annual rainfall between 1750 and 2000 mm per year and an average temperature of 18.6 °C. This catchment is representative of the basaltic plateau region of the Serra Geral Formation, where the main soil classes found are Ferralsols/*Latossolos* (80 %), Nitisols/*Nitossolos* (18 %) and Acrisols/*Argissolos* (2 %) (IUSS Working Group WRB, 2015/Santos et al., 2013), with mineralogy dominated by iron oxides and kaolinite. The main land-use is cropland (89 %), mainly cultivated under no-tillage with soybeans (*Glycine max*) in summer and wheat (*Triticum aestivum*) for grain production, oats (*Avena sativa* and *Avena strigosa*) and ryegrass (*Lolium multiflorum*) for feeding dairy cattle or used as a cover crop to protect the soil during winter. Improved soil management practices, such as no-till farming, have significantly contributed to notable reductions in erosion rates compared to the 1980s. Despite these advancements, the region still experiences high erosion rates due to the poor quality of the no-till system carried out in the region (Didoné et al., 2019).

The relief of the river catchment is characterized by gentle slopes (6-9 %) at the highest positions of the landscape and steeper slopes (10-14 %) near the drainage channels, with altitudes ranging from 270 to 480 m. The catchment outlet is located next to the monitoring point number 75,200,000 of the National Water Agency (ANA) (28° 27′ 22″ S, 53° 58′ 24″ O) in the municipality of Ijuí.

Source and sediment sampling

Soil composite samples (n = 181) were collected in representative sites to characterize the four potential sediment sources, including cropland (n = 78), stream bank (n = 36), unpaved road (n = 40) and pasture (n = 27). Samples were collected from the topsoil layer (0.00-0.05 m) of surface sources (cropland and pasture) (Figure 2). In stream banks, the samples were collected in the exposed sidewall, avoiding the material from the uppermost layer. Samples of unpaved roads were taken mainly in the roadsides where erosion is the most evident, which corresponds to the deeper layers of the soil. Moreover, this strategy avoids sampling material transiting from other sources before reaching the river network. For all source samples, around ten sub-samples were collected within a radius of approximately 50 m and well-mixed to prepare a composite sample representative of the area. Samples were collected at sites sensitive to erosion and connected to the stream network. A total of 44 sediment samples were collected from March 2011 to March 2013, including suspended sediment samples collected by time

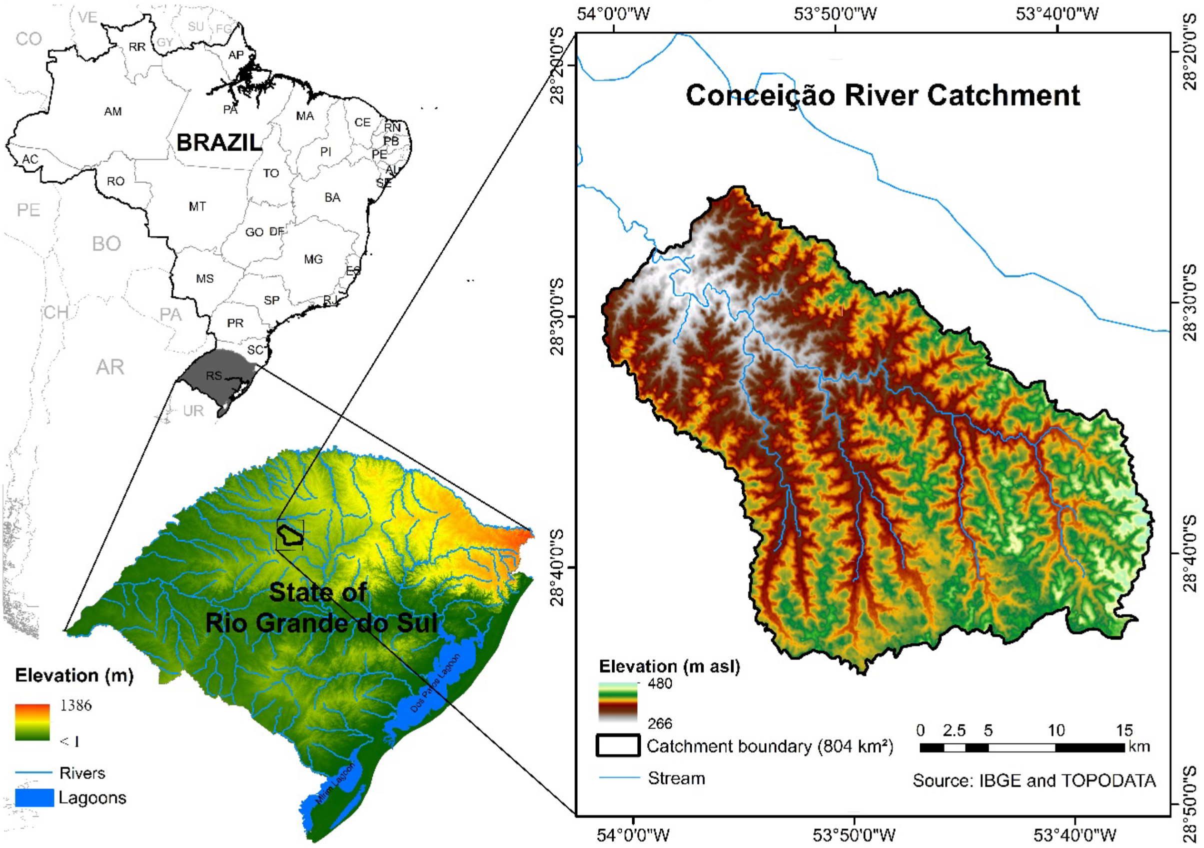

Figure 1. Location of the Conceição River catchment in Southern Brazil and digital elevation model.

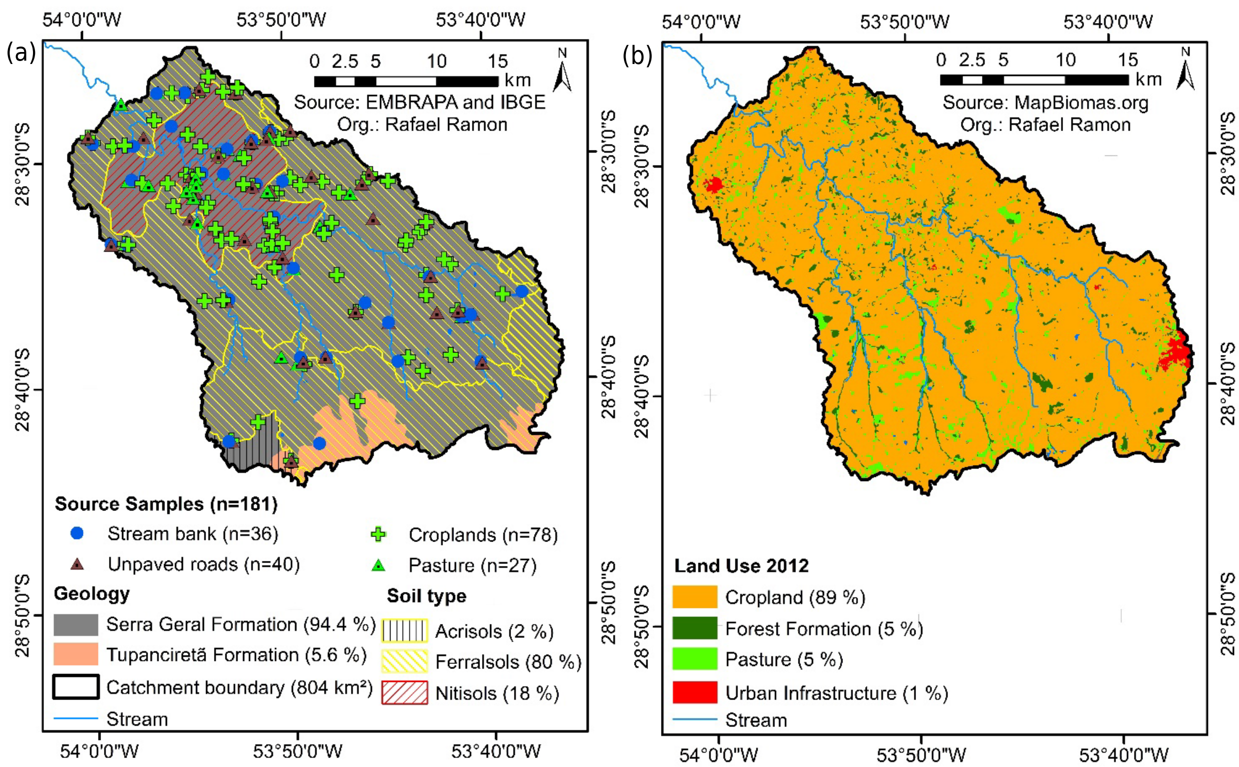

Figure 2. Lithological formations, soil types and source sample location (a); and land-use map for the year 2012 (b) in the Conceição River catchment, Southern Brazil.

integrating suspended sediment samplers (TISS) (n = 8), fine sediment deposited on the riverbed (n = 15) and suspended sediment samples collected in the water column during storm events (n = 21). Source and sediment samples were oven-dried at 50 °C, gently disaggregated using a pestle and mortar and dry-sieved to 63 µm to avoid the particle size effect (Laceby et al., 2017).

Artificial mixtures

Equal proportions of the samples from each sediment source were mixed in the laboratory to prepare a single reference sample for each corresponding source (cropland, pasture, unpaved roads, and stream bank). Subsequently, these four reference samples were mixed, combining 97 different proportions from each source. These mixtures were then used to calibrate the multivariate mathematical models used to estimate the respective source contributions to the sediment samples.

Spectral analyses

Diffuse reflectance spectral analyses were carried out in the ultraviolet–visible (UV-VIS, 200–800 nm with 1 nm step), near-infrared (NIR, 1000-2500 nm with 1 nm step) and mid-infrared wavelengths (MIR, 2500–25,000 nm with 5 nm step). The UV-VIS spectra were measured for each powdered sample using a Cary 5000 UV-VIS-NIR spectrophotometer (Varian, Palo Alto, CA, USA) at room temperature, using $BaSO_4$ as a 100 % reflectance standard. The NIR spectra was measured using a Nicolet 26,700 FTIR spectrometer (Waltham, Massachusetts, USA) in diffuse reflectance mode with an integrating sphere and a InGaAs detector with 100 readings per spectrum. The MIR spectra were measured with a Nicolet 510-FTIR (Thermo Electron Scientific, Madison, WI, USA) spectrometer in diffuse reflectance mode with 100 readings per spectrum. For MIR analysis, a direct current of air was used (dry and without CO_2) to eliminate CO_2 and water from the spectrometer in order not to interfere with scanning when obtaining the spectra.

Spectral pre-processing techniques and multivariate model calibration and validation

After the spectra acquisition, they were submitted to three different spectral pre-processing techniques to outline features of interest and reduce the physical variability due to light dispersion or systematic variations due to environmental and instrumental conditions (Barnes et al., 1989; Dotto et al., 2019). Standard Normal Variate (SNV), Detrend (DET) and Savitzky-Golay Derivate (SGD) pre-processing techniques were then compared to the raw spectra (RAW). These pre-processing techniques were chosen based on the performance of previous studies available in the literature, such as found in Tiecher et al. (2021). The SNV normalize the spectral data to correct for light scatter. The DET allows the correction of wavelength-dependent scattering effects, normalizing the spectral data. The SGD reduces the high-frequency noise in the signal by smoothing properties and reducing the low-frequency signal by differentiation. The SGD was applied with a first derivative using a first-order polynomial and an 11-nm search window, where the latter window search was defined based on prior testing.

To develop the prediction models, a parametric multivariate model, the PLSR (method ‘pls’) (R pls package) (Mevik et al., 2016), and a non-parametric model, the SVM (method ‘svmLinear’) (R e1071 package) (Meyer et al., 2019), were used. The PLSR are frequently used because of their simplicity, robustness, high performance and easy accessibility. In this method, the raw spectra are converted into latent variables to reduce the dimensionality, and multiple linear regressions are then implemented between the measured spectra and the variable of interest. The SVM has the advantage of being able to model non-linear relations, since the latter are expected between spectral variables and organo-mineral components of the soil (Tiecher et al., 2021; Viscarra Rossel and Behrens, 2010). More detailed information about the process utilized by these models to estimate source contributions can be found in Tiecher et al. (2021) and Naibo et al. (2022). The models were calibrated and validated by cross-validation with 10 k-fold, using a random division into segments. The performance of the models in estimating the contribution from each source to sediment was evaluated by the following quantification statistics of accuracy: coefficient of determination (R^2) (Equation 1) and root mean square error of prediction (RMSE) (Equation 2).

$$R^2 = \frac{\sum_{i=1}^{N}\left(\hat{y}_i - \bar{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}_i\right)^2} \qquad \text{Eq. 1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \qquad \text{Eq. 2}$$

in which: $\hat{y}$ is the predicted value, $\bar{y}$ is the observed mean value, y is the observed value, and n is the number of samples.

Moreover, the model performance was also compared based on the prediction results of source contributions to sediment samples. For this purpose, as suggested by Legout et al. (2013) and Tiecher et al. (2021), we considered that the closer the total sum of the source contributions in sediment samples is from 100 %, the higher is the model performance. Sums of contributions from each source that differ from 100 % are a consequence of this approach, in which a model is calibrated and validated to predict the contribution of a single source based on artificial mixtures. Unlike mixing models, in which, through interactions, the model adjusts the contribution of each source so that the sum is 100 %, in this approach, this interaction between models for each source does not occur. All pre-processing and data modeling was performed in R software (R Development Core Team, 2020). The statistical performance was compared with an ANOVA at a 5 % significance level. When significant, the difference between the means of the multivariate model, the pre-processing technique and the spectral range combination means were evaluated by the Tukey test ($p<0.05$).

Building spectroscopy models for different sediment sources

Spectroscopic models were built considering three approaches. First, the models were calibrated for each of the four potential sources considered (cropland, pasture, unpaved roads and stream bank). Previous studies conducted in the same catchment already showed a high similarity of the soil properties under cropland and pasture (Tiecher et al., 2018; Ramon et al., 2020). For this reason and to avoid misclassification between these two surface/agricultural sources, they were grouped as surface sources in a second approach. Merging sources with similar characteristics is a method that has already been used in previous studies to have greater reliability in model prediction (Poulenard et al., 2009). In addition, a third approach was tested grouping unpaved roads and stream banks into a single subsurface source category. Accordingly, in that case, only two sources (i.e., surface and subsurface) were considered.

Regardless of the source considered, in total, 56 models were calibrated for each sediment source with the combination of two multivariate models (PLSR, and SVM), four spectra pre-processing techniques (RAW, SNV, DET, and SGD) and seven spectral ranges (UV-VIS, NIR, MIR, UV-VIS+NIR, UV-VIS+MIR, NIR+MIR, UV-VIS+NIR+MIR). Therefore, altogether, 504 models were constructed in the current research, considering four (n = 224), three (n = 168) and two (n = 112) potential sources, respectively.

RESULTS AND DISCUSSION

Effect of multivariate model calibration

The three approaches, with the combinations of multivariate models, spectral ranges and pre-processing techniques, resulted in 504 models. Overall, the SVM provided more accurate models compared to PLSR, as indicated by the higher R^2 and lower RMSE values obtained with SVM, regardless the number of sediment sources considered. Table 1 shows the statistics obtained for each combination of multivariate models, pre-processing technique and spectral range. The PLSR provided the lowest RMSE and the highest R^2 individual values among the three approaches. However, R^2 and RMSE varied widely depending on the source, the pre-processing technique and the spectral range considered. In contrast, the SVM model provided a lower variation in the associated R^2 and RMSE values, showing a better performance regardless of the spectral range or pre-processing technique considered (average values of each approach: R^2 >0.97 and RMSE <4.14 %) (Table 1). The SVM model managed to establish mathematical relationships to express non-linear correlations between the organo-mineral composition and the spectral behavior of sediment and soil samples (Viscarra Rossel and Behrens, 2010; Wijewardane et al., 2016). Due to the ability and flexibility to handle non-linear relationships between spectra and soil/sediment properties, the SVM has been more appropriate for calibration of large heterogeneous samples, providing better results than PLSR models (Araújo et al., 2014; Lucà et al., 2017; Tiecher et al., 2021). Based on the general statistical performance of the multivariate model calibration, the SVM was considered the most robust model for predicting sediment source contributions. The PLSR model also provided satisfactory validation statistical values, with R^2 values close to one and low RMSE values, as already observed in other studies (Legout et al., 2013; Ni et al., 2019). Since there is no consensus in the literature on how to define the best prediction models and the best pre-processing method, preliminary tests considering all the possible combinations are recommended, in particular in large and complex catchments with a higher number of potential sources (Tiecher et al., 2021).

Effect of pre-processing techniques

Among the tested pre-processing techniques, the SGD showed the lowest RMSE and the highest R^2 values for the SVM model (Table 1). Although the three pre-processing techniques (SGD, SNV and DET) resulted in models with similar accuracies, only the SGD technique was significantly better than the RAW spectra. Based on this result, the SGD was considered the best pre-processing technique for SVM models. The opposite situation was observed for the PLSR model, in which the SGD pre-processing technique resulted in the highest RMSE and the lowest R^2 values, even when compared to the RAW spectra (without spectra pre-processing). With the PLSR models, the SNV resulted in the lowest RMSE values. However, it did not differ significantly from the RAW spectra in all the approaches and from other pre-processing techniques for the approaches considering four and two sources, respectively (Table 1). The SNV was significantly different from the SGD pre-processing technique in the approach with three sources. Tiecher et al. (2021) suggested that pre-processing techniques may potentially improve model calibration in larger catchments with soils showing homogeneous mineralogical compositions. However, this was not observed in the current research, since the pre-processing techniques did not result in major gains of model calibration. According to these results, the SGD pre-processing technique is recommended as an alternative to improve SVM model calibration. In contrast, for the PLSR model, no pre-processing is recommended, since none of the tested methods led to significant improvements in the statistical parameters.

Effect of spectral ranges

In all three approaches, with four, three and two potential sediment sources, the models calibrated with the UV-VIS spectra resulted in higher RMSE and lower R^2 values, regardless of the pre-processing or multivariate model used (Table 1). The other spectral ranges, including the UV-VIS spectra (especially UV-VIS+NIR and UV-VIS+MIR) also resulted in low accuracy models, mainly when combined with the PLSR approach (Table 1). The calibration of the SVM model provided better results than the PLSR model for UV-VIS spectra and better results than those obtained with most of the other spectral ranges.

Table 1. Comparison of spectroscopic model accuracy indicated by R^2 and RMSE as affected by pre-processing technique and spectral range using PLSR and SVM methods considering four, three and two potential sources of sediments

Parameter	Combination of pre-processing technique and spectral range		Four sediment sources (cropland, pasture, unpaved roads, stream bank)				Three sediment sources (surface, unpaved roads, stream bank)				Two sediment sources (surface, subsurface)			
			PLSR		SVM		PLSR		SVM		PLSR		SVM	
RMSE	Pre-processing	RAW	5.26	a	4.83	a	5.29	ab	4.44	a	6.92	a	5.13	a
		SNV	5.62	a	4.27	a	4.66	b	3.96	ab	5.96	a	4.40	a
		DET	5.48	a	4.10	ab	5.32	ab	3.83	ab	6.92	a	4.21	ab
		SGD	6.86	a	3.10	b	6.85	a	2.93	b	7.52	a	3.38	b
	Spectral range	UV	17.77	a	12.4	a	15.89	a	10.04	a	21.27	a	11.8	a
		NIR	2.92	bc	2.56	b	2.29	c	2.58	b	2.39	d	2.88	b
		MIR	4.17	bc	2.74	b	4.48	bc	2.84	b	6.17	bc	3.18	b
		UV+NIR	5.16	b	2.61	b	4.68	bc	2.61	b	3.84	cd	2.77	b
		UV+MIR	5.40	b	2.80	b	6.22	b	2.79	b	7.99	b	3.21	b
		NIR+MIR	1.88	c	2.68	b	1.60	c	2.79	b	1.74	d	3.15	b
		UV+NIR +MIR	3.33	bc	2.70	b	3.55	bc	2.79	b	4.40	cd	3.13	b
	Mean		5.80	Aa	4.07	Ab	5.53	Aa	3.79	Ab	6.83	Aa	4.56	Ab
R^2	Pre-processing	RAW	0.94	a	0.94	b	0.95	ab	0.96	b	0.93	a	0.96	b
		SNV	0.94	a	0.96	ab	0.96	a	0.97	ab	0.95	a	0.96	ab
		DET	0.93	a	0.97	ab	0.95	ab	0.98	ab	0.92	a	0.98	ab
		SGD	0.90	a	0.99	a	0.91	b	0.99	a	0.89	a	0.99	a
	Spectral range	UV	0.61	b	0.80	b	0.73	b	0.88	b	0.60	b	0.87	b
		NIR	0.99	a	0.99	a	0.99	a	0.99	a	0.99	a	0.99	a
		MIR	0.98	a	0.99	a	0.98	a	0.99	a	0.96	a	0.99	a
		UV+NIR	0.96	a	0.99	a	0.96	a	0.99	a	0.99	a	0.99	a
		UV+MIR	0.96	a	0.98	a	0.94	a	0.99	a	0.94	a	0.99	a
		NIR+MIR	1.00	a	0.99	a	1.00	a	0.99	a	1.00	a	0.99	a
		UV+NIR +MIR	0.98	a	0.99	a	0.98	a	0.99	a	0.98	a	0.99	a
	Mean		0.93	Ab	0.96	Aa	0.94	Ab	0.97	Aa	0.92	Ab	0.97	Aa

Means followed by the same low case letters in the column, comparing pre-processing techniques and spectral ranges, are not significantly different according to the Tukey test at $p<0.05$. Means followed by the same capital letters in the raw, comparing the number of sediment sources considered, are not significantly different according to the Tukey test at $p<0.05$. Means followed by the same low case letters in the raw, comparing multivariate models, are not significantly different according to the Tukey test at $p<0.05$. PLSR: Partial Least Square Regression; SVM: Support Vector Machine; UV: Ultra-violet-visible; NIR: Near Infrared; MIR: Mid Infrared. In the "Pre-processing" rows, values indicated are the average of all Spectral range and sources, while in the "Spectral range" rows, values are the average of all pre-processing and sources.

With the exception of the UV-VIS spectra, the other spectral ranges resulted in R^2 values close to one. For the PLSR, the difference in model performance between the spectral ranges considered was higher, especially when comparing the RMSE values. The best results were observed for the combination of NIR+MIR, which differs statistically in all the approaches from the UV-VIS, UV-VIS+NIR and UV-VIS+MIR spectral ranges. The UV-VIS range is highly influenced by the content of iron oxide in soils and sediments (Viscarra Rossel and Behrens, 2010). In the Conceição River catchment, the occurrence of highly weathered and homogeneous soils with the absence of clear composition difference between soil layers and land-use sources, had already been demonstrated with other tracers (geochemical composition, magnetic and colour parameters), which complexifies their discrimination (Tiecher et al., 2018; Ramon et al., 2020). For similar reasons, differences are likely not sufficient between the considered sources to provide discrimination in the UV-VIS range.

Absorption features in NIR and MIR wavelengths are mainly related to the content in structural water, clay minerals and organic matter, and they can provide a large quantity of information about the sample composition (Viscarra Rossel and Behrens, 2010; Tiecher et al., 2017). Differences in organic carbon content between surface (22.0 to 25.9 g kg^{-1}) and subsurface (7.4 to 15.7 g kg^{-1}) layers (Ramon et al., 2020) may justify the better discrimination obtained in the approach with two sources (surface and subsurface), which presented a lower uncertainty associated with the discrimination of the sources (11.9 %) and a better percentage of correctly classified source samples (92.8 %) compared to the three-source approach (Tables 2, 3, and 4). Indeed, the NIR and MIR spectra have been widely and successfully used to predict soil organic carbon contents (Knox et al., 2015; Moura-Bueno et al., 2019; Ng et al., 2019) and, further, they may provide good discrimination between sources with different soil organic carbon contents. In addition, the MIR spectra have been proven to be efficient in predicting different soil organic carbon fractions (Reeves III et al., 2006; Bellon-Maurel and McBratney, 2011), providing further information about the organic composition of the samples that may improve the discrimination between sources.

Table 2. Discriminant function analysis outputs for the approach with four sources in the Conceição River catchment, Southern Brazil. This analysis was made after the application of the SGD pre-processing technique

Selected PCA components	**1, 15, 3, 4, 5, 6, 7, 14, 9, 10, 19**
Wilk's Lambda	0.1068
Variance explained by the variables (%)	89.32
Squared Mahalanobis distances	
Cropland vs Pasture	1.98
Cropland vs Stream Bank	6.41
Cropland vs Unpaved Road	21.40
Pasture vs Stream Bank	6.65
Pasture vs Unpaved Road	28.36
Stream Bank vs Unpaved Road	15.79
Average	13.43
p-levels	
Cropland vs Pasture	<0.001
Cropland vs Stream Bank	<0.001
Cropland vs Unpaved Road	<0.001
Pasture vs Stream Bank	<0.001
Pasture vs Unpaved Road	<0.001
Stream Bank vs Unpaved Road	<0.001
Correctly classified source samples **(%)**	
Croplands	96.15
Pasture	40.74
Stream Bank	77.78
Unpaved Road	87.5
Average	82.3
Uncertainty associated with the discrimination of the source **(%)**	
Cropland	21.54
Pasture	55.38
Stream Bank	30.58
Unpaved Road	14.14
Average	30.41

Table 3. Discriminant function analysis outputs for the approach with three sources in the Conceição River catchment, Southern Brazil. This analysis was made after the application of the SGD pre-processing technique

Selected PCA components	1, 14, 3, 4, 5, 6, 7, 19, 9
Wilk's Lambda	0.1369
Variance explained by the variables (%)	86.31
Squared Mahalanobis distances	
Surface vs Stream Bank	5.83
Surface vs Unpaved Road	21.44
Stream Bank vs Unpaved Road	14.30
Average	13.86
p-levels	
Surface vs Stream Bank	<0.001
Surface vs Unpaved Road	<0.001
Stream Bank vs Unpaved Road	<0.001
Correctly classified source samples **(%)**	
Surface	98.1
Stream Bank	77.78
Unpaved Road	82.5
Average	90.61
Uncertainty associated with the discrimination of the source **(%)**	
Surface	6.48
Stream Bank	29.42
Unpaved Road	15.64
Average	17.18

All models were well-calibrated, except for the models calibrated with the UV-VIS spectra alone, showing a R^2 above 0.94 and an RMSE below 7.99 % (Table 1). This results are better than those observed for another small catchment in southern Brazil, where mean validation RMSE values for all pre-processing techniques and spectral ranges exceeded 7.4 % for PLSR and SVM models (Tiecher et al., 2021). In the current research, the mean RMSE values remained below 6.83 % for the PLSR models and 4.56 % for the SVM models (Table 1).

Despite the satisfactory statistical values obtained for the model calibrations (Figure 3, as an example), the sum of individual source contribution predictions for each sediment sample was not always equal to 100 % when using multivariate models. This type of result differs from those of the conventional mixing models that require the prediction of individual source contributions to lie in a range comprised between 0 and 100 %, and that the sum of multiple sources equals 100 %. According to the literature, when unconstrained, the closer this sum is to 100 %, the better is the performance of the models (Legout et al., 2013; Tiecher et al., 2021). Table 5 shows the average absolute sums for all sediment samples and for each spectral range considered. The results showed the UV-VIS spectra alone provide the least satisfactory results, differing statistically from those obtained with other spectral ranges regardless of the number of sources considered, and the multivariate models used. Observing the PLSR model calibrated with UV-VIS spectra, after the SGD pre-processing, the predicted contributions were underestimated compared to the observed proportions, in particular for the model adjusted for cropland, pasture, and unpaved road sources (data not showed). This underestimation led to a lower contribution predicted for each of these sources and, consequently, a lower absolute sum, resulting in values closer to 100 %. In addition, a statistical difference between spectral ranges was also observed in the approach with three sources for the PLSR model, in which the

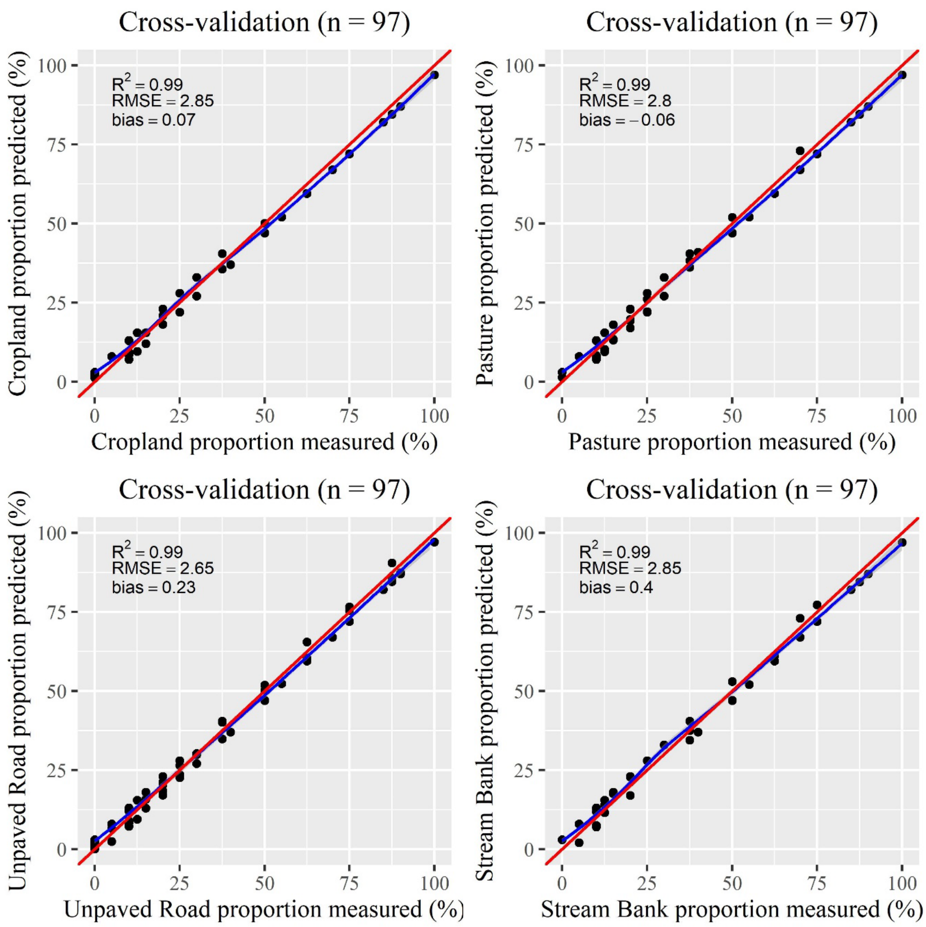

Figure 3. Support vector machine model calibration and validation with the MIR spectra after SGD pre-processing.

best quality result was observed for the combination for UV-VIS+NIR+MIR (142.1 %). This analysis contributed to the identification of the best spectral range to use for source tracing. However, from these results, the unique conclusion that can be drawn is the use of UV-VIS spectra requires more caution when they are considered potential tracers of sediment sources.

Effect of the number of sediment sources considered

Table 1 presents the average statistical values for each multivariate model calibration and the number of sources considered. Figures 4 and 5 show the Mahalanobis distances between source samples, and they illustrate that soil sample properties from cropland and pasture overlap. The combination of these two sources in the third approach increased the number of samples for the model calibration, and due to their similarity, this combined approach may result in better statistical results. The opposite situation is observed when combining unpaved roads and stream banks. Although they both consist of subsurface material, they are clearly more different from each other compared to the two types of surface sources. The increase in within-source variability when grouping sources together may reduce the quality of model calibration. Previous studies have already warned about the uncertainty of the method application in more complex catchments with variations in soil types and land-uses (Poulenard et al., 2012).

No significant difference was observed when the average statistical values for the three approaches are compared (Table 1). However, a higher uncertainty in the model predictions is expected with the increase in the number of sources (Poulenard et al., 2009; Evrard et al., 2022). Since no significant difference was observed between

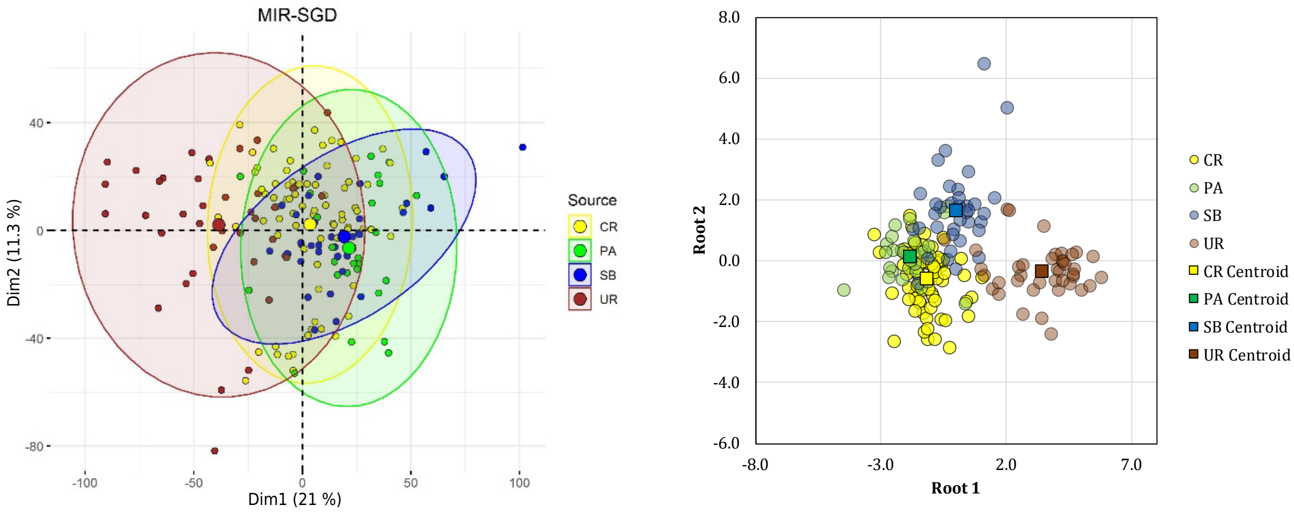

Figure 4. Principal component analysis biplots with four sources for the SVM model calibrated with the combination of MIR spectral range and SGD pre-processing technique (left). Two-dimensional scatter plot of the first and second functions derived from the DFA (right). CR: cropland; PA: pasture; SB: stream bank; and UR: unpaved road.

the tested approaches and a better discrimination is expected when reducing the source number, a principal component analysis (PCA) was applied to the soil data to obtain the natural clustering of each source for the three approaches. Subsequently, a discriminant function analysis (DFA) was applied to the twenty-first PCA components to verify the discrimination capability between sources. The DFA was applied to analyze the impact of reducing the number of sources on the discrimination power. The percentage of correctly classified samples by the DFA increased with the reduction in source number, from 83.4 to 90.6 % from four to two sources, respectively (Tables 2, 3, and 4). The uncertainty associated with the discrimination of the sources was also reduced, from 28.8 to 15.6 %. This is illustrated in figures 6 and 7, and this result further demonstrates the reduction of the source number minimizes the uncertainty in the model predictions.

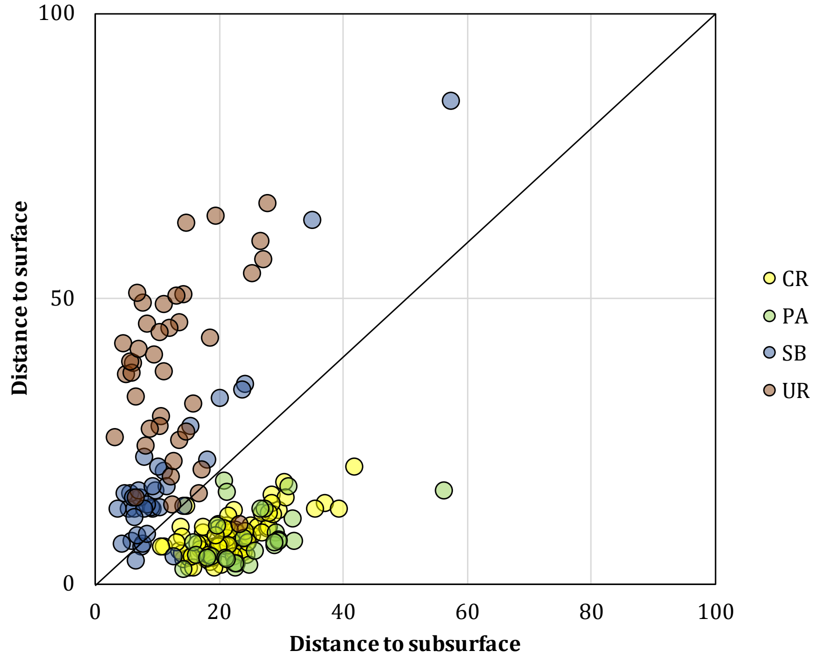

Figure 5. Mahalanobis distance from the discriminant analysis to differentiate subsurface samples from surface samples in the Conceição River catchment, Southern Brazil. CR: cropland; PA: pasture; SB: stream bank; and UR: unpaved road.

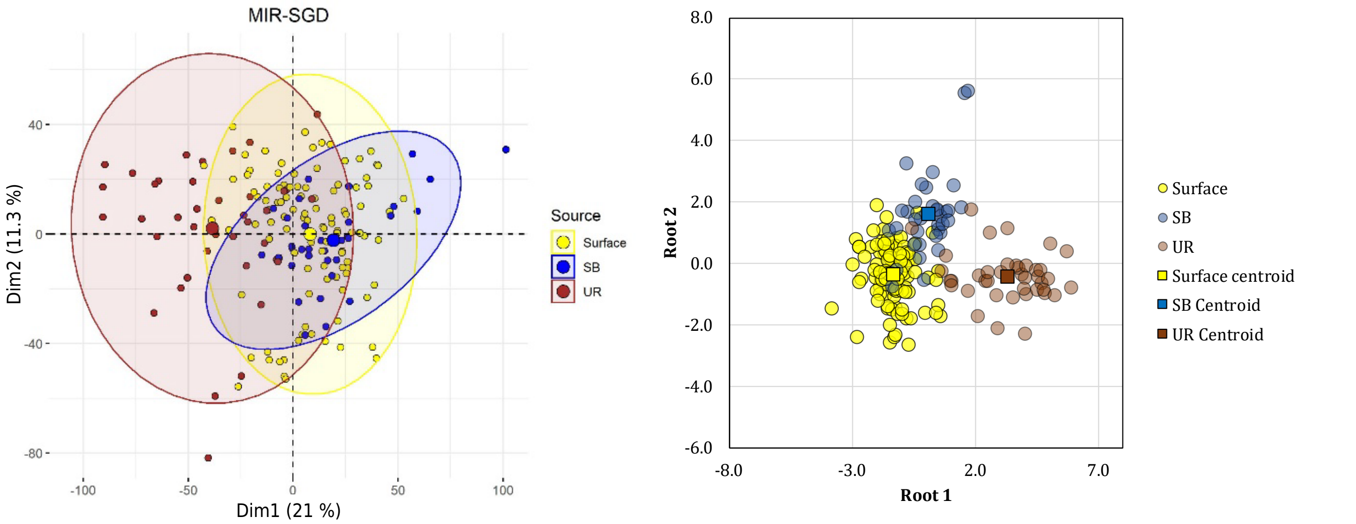

Figure 6. Principal component analysis biplots with three sources for the SVM model calibrated with the combination of MIR spectral range and SGD pre-processing technique (left). Two-dimensional scatter plot of the first and second functions derived from the DFA (right). Surface: surface sources (cropland+pasture); SB: stream bank; and UR: unpaved road.

Although the DFA results were obtained from the MIR spectra after SGD pre-processing, it can be seen from figures 8, 9 and 10 that the reduction in source numbers decreases the differences between all 56 models with the combination of multivariate models, pre-processing and spectral range in the prediction of the source contributions. In the approach with two sources, for almost all model combinations, there was no significant difference between them (Figure 8). Only five models, the SVM model obtained from the UV-VIS spectra alone and the PLSR model calibrated with UV-VIS spectra after SNV pre-processing, differed statistically from all the others. For the approach with three sources, nine models were statistically different and provided lower-quality results than

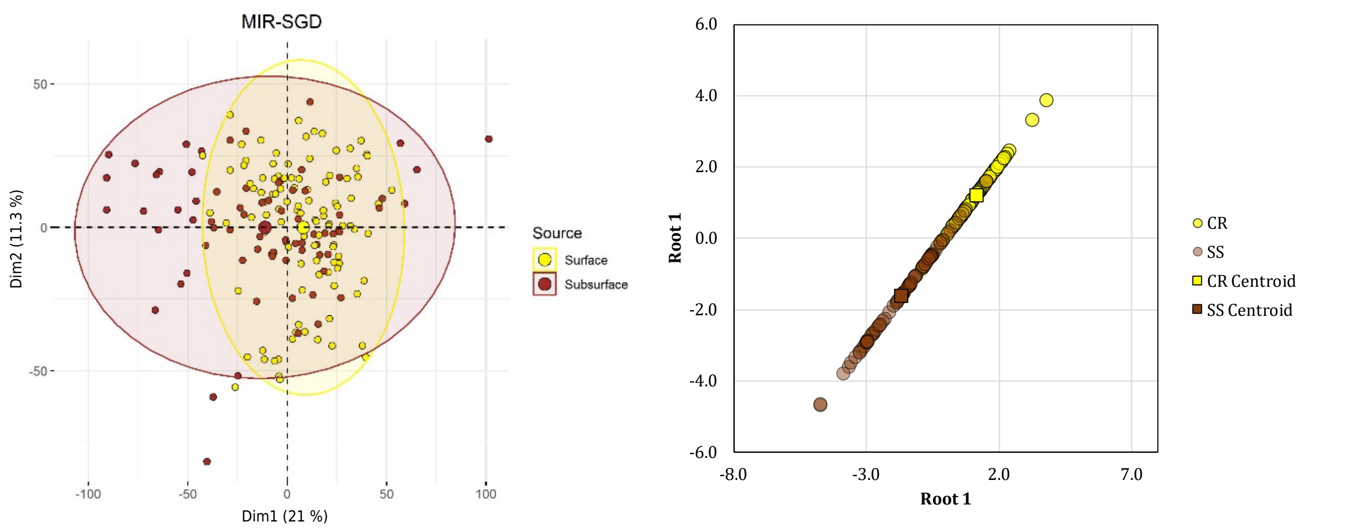

Figure 7. Principal component analysis biplots with two sources (left). Two-dimensional scatter plot of the first and second functions derived from the DFA (right). Surface: surface sources (cropland+pasture); Subsurface: subsurface sources (stream bank and unpaved road).

Table 4. Discriminant function analysis outputs for the approach with two sources in the Conceição River catchment, Southern Brazil. This analysis was made after the application of the SGD pre-processing technique

Selected PCA components	**1, 10, 3, 4, 5, 6, 7, 14, 9**
Wilk's Lambda	0.34
Variance explained by the variables (%)	66.27
Squared Mahalanobis distances	
Surface vs Subsurface	7.97
p-levels	
Surface vs Subsurface	<0.001
Correctly classified source samples **(%)**	
Surface	97.14
Subsurface	86.84
Average	92.82
Uncertainty associated with the discrimination of the source **(%)**	
Surface	5.64
Subsurface	18.16
Average	11.90

the best set, while for the approach with four sources, sixteen model combinations were found in that situation. As the number of sources is reduced, the mean of the absolute sum of individual source contributions becomes closer to 100 % and the effect from the pre-processing technique, spectral range and multivariate model is expected to decrease.

Sediment source contributions

According to the statistical analysis already discussed in the previous section, the combination SVM-SGD-MIR provided the best suite to predict sediment source contributions. Therefore, the sediment source contributions predicted by this combination are presented here for the approaches with different numbers of potential sources. A higher uncertainty was observed in the first approach with four sources, with source contributions exceeding 100 % and including negative contributions predicted for cropland in particular (Figure 11). The low discrimination between pasture and cropland samples (Figure 4) indicates the occurrence of limited differences in the soil properties between these two superficial sources. Consequently, the model calibration for these two sources may result in unreliable

Table 5. Comparison of the absolute sum of source contributions to individual sediment samples affected by spectral range using PLSR and SVM methods considering four, three and two potential sources of sediments, in the Conceição River catchment, Southern Brazil

Spectral range	**Four sediment sources (cropland, pasture, unpaved roads, stream bank)**				**Three sediment sources (surface, unpaved roads, stream bank)**				**Two sediment sources (surface, subsurface)**			
	PLSR		**SVM**		**PLSR**		**SVM**		**PLSR**		**SVM**	
	%											
UV	284.3	b	528.7	a	285.1	a	536.4	a	230.0	a	486.9	a
NIR	280.4	b	288.1	b	180.3	bc	181.9	b	152.1	b	150.5	b
MIR	228.1	bc	203.5	d	150.0	bc	152.7	b	127.8	b	133.3	b
UV+NIR	294.1	a	318.3	b	184.7	b	204.7	b	149.9	b	172.3	b
UV+MIR	211.7	c	208.7	cd	157.1	bc	158.5	b	146.1	b	129.2	b
NIR+MIR	333.0	a	309.8	b	165.0	bc	190.2	b	161.5	b	150.5	b
UV+NIR+MIR	271.8	abc	279.6	bc	142.1	c	178.5	b	128.0	b	140.7	b

Means followed by the same letters in the column, comparing spectral ranges, are not significantly different according to the Tukey test at $p<0.05$. PLSR: Partial Least Square Regression; SVM: Support Vector Machine; UV: Ultra-violet-visible; NIR: Near Infrared; MIR: Mid Infrared.

predictions due to their similar properties, even with the optimum fit of the models for the two sources (Figure 3). The results show an underestimation of cropland contributions with negative values, while pastures displayed higher contributions, especially for Event and TISS samples, where, in some cases, their contribution exceeded 100 % (Figure 11).

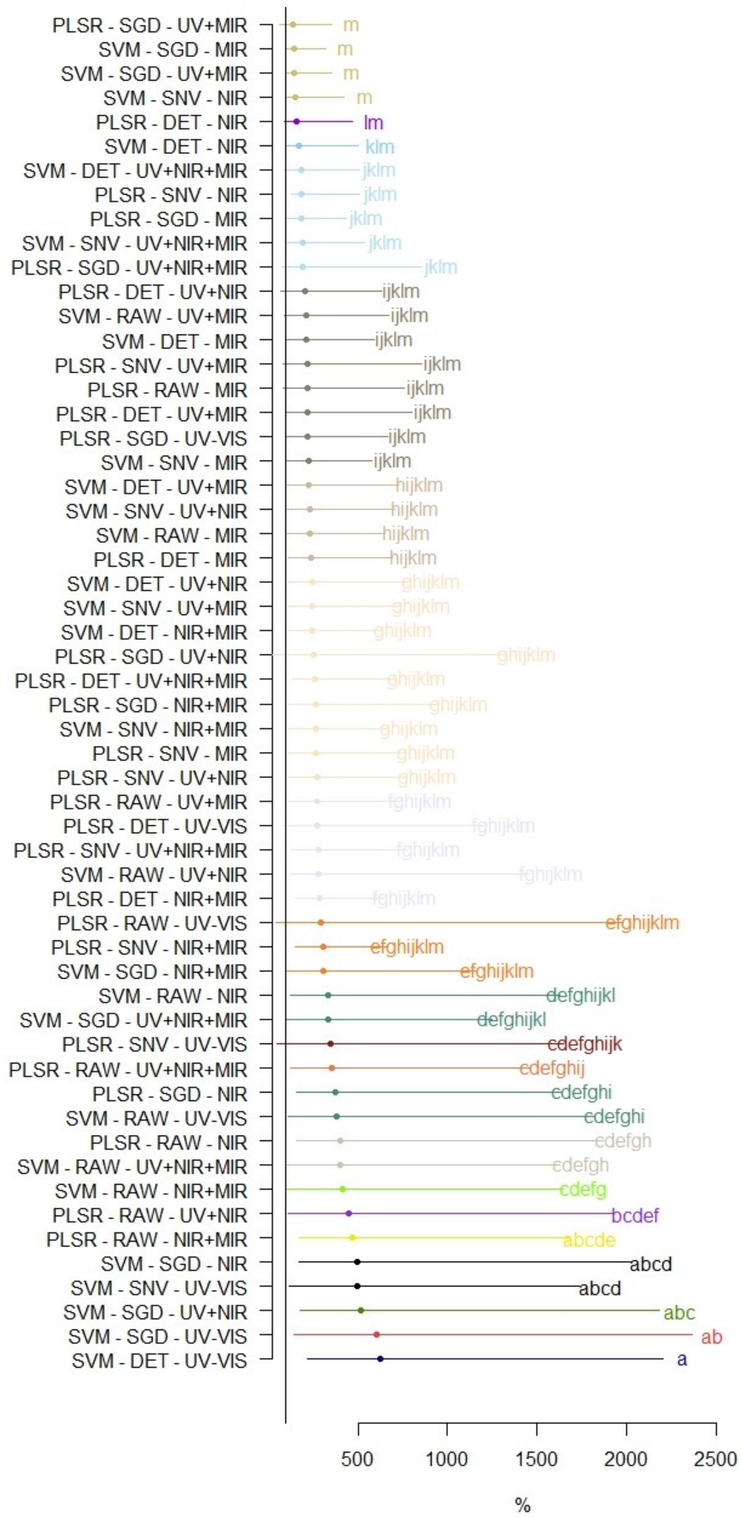

Figure 8. Mean test of the source contribution absolute sum average of the sediment samples for the approach with four sources in the Conceição River catchment, Southern Brazil. The vertical black line indicates the 100 %. Means followed by the same letters are not significantly different according to the Tukey test at p<0.05. Statistically equal means are grouped by colour. UV or UV-VIS: ultraviolet-visible wavelength.

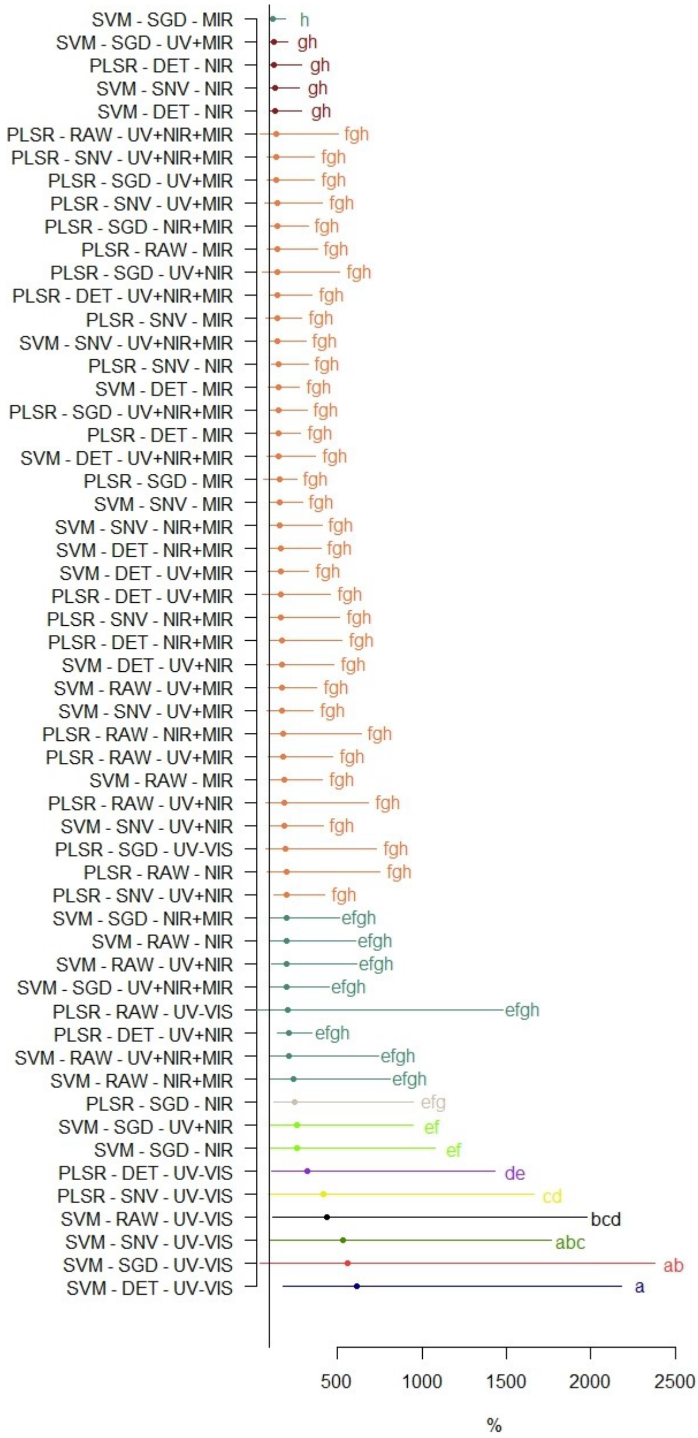

Figure 9. Mean test of the source contribution absolute sum average of the sediment samples for the approach with three sources in the Conceição River catchment, Southern Brazil. Vertical black line indicates the 100 %. Means followed by the same letters are not significantly different according to the Tukey test at p<0.05. Statistically equal means are grouped by colour. UV or UV-VIS: ultraviolet-visible wavelength.

Differences in source contributions were observed according to the types of sediment samples. Pasture provided the main sources of suspended sediment during rainfall events and also for the TISS samples (Figure 11). However, considering the low discrimination between cropland and pasture sources, it can be assumed that the main sediment sources for Event and TISS samples are surface sources, as indicated in figure 12. This result agrees with those obtained by previous studies conducted in this catchment (Tiecher et al., 2018; Ramon et al., 2020). In addition, the absence of runoff control, appropriate soil

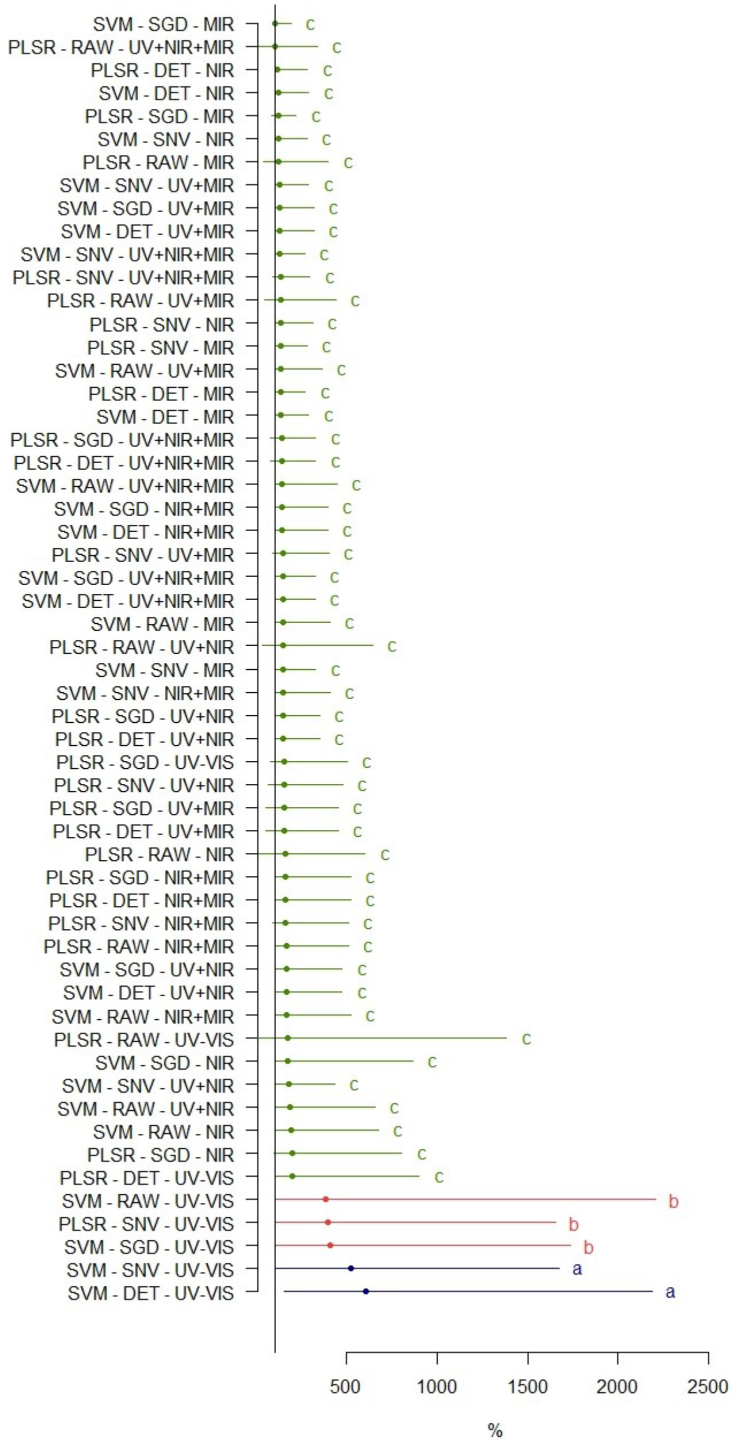

Figure 10. Mean test of the source contribution absolute sum average of the sediment samples for the approach with two sources in the Conceição River catchment, Southern Brazil. Vertical black line indicates the 100 %. Means followed by the same letters are not significantly different according to the Tukey test at $p<0.05$. Statistically equal means are grouped by colour. UV or UV-VIS; ultraviolet-visible wavelength.

conservation practices and the high erosivity potential of rainfall events in the region, may generate significant transfers of suspended sediment (Didoné et al., 2017). Accordingly, surface sources supplied a greater contribution to sediment samples collected during events as well as for the TISS material. It is usually during the flood events that most of the sediment fluxes take place (Minella et al., 2017), which indicates that large amounts of sediment and associated nutrients may be transported from agricultural areas and delivered to water bodies, requiring interventions to avoid soil and nutrient losses from

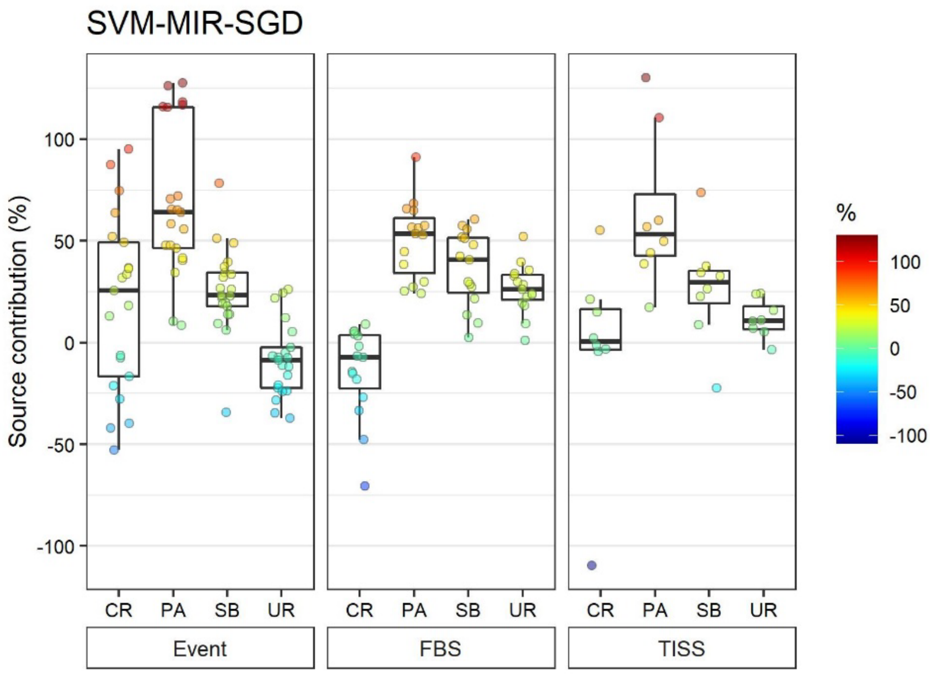

Figure 11. Sediment source contributions according to the sampling method predicted from the SVM-SGD-MIR models in the Conceição River catchment, Southern Brazil. Event: Rainfall runoff event; FBS: Fine bed sediment samples; TISS: Time integrated suspended sediment samples; CR: cropland; PA: pasture; SB: stream bank; and UR: unpaved road.

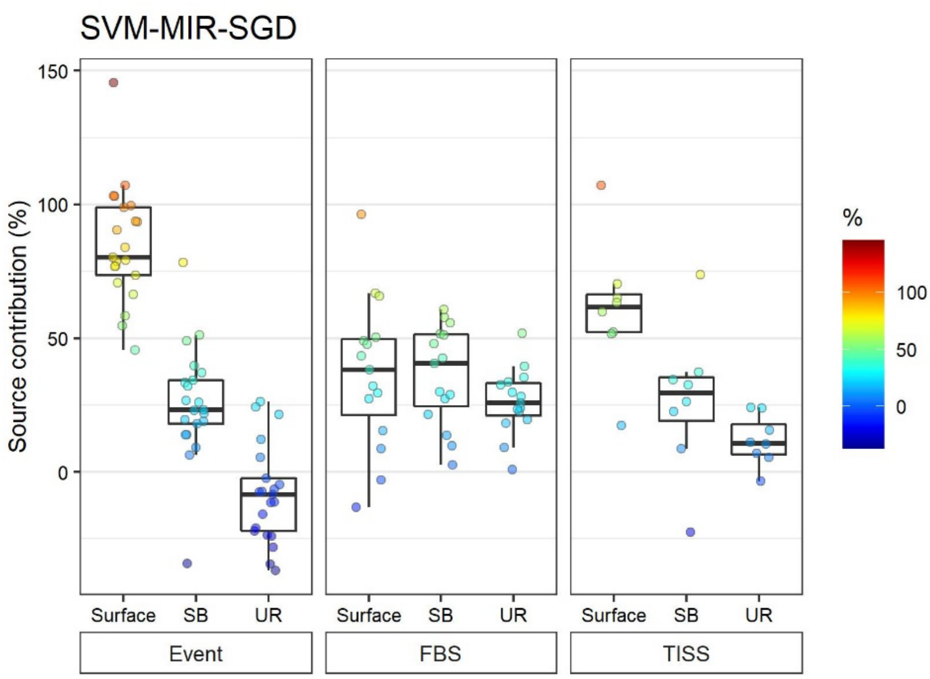

Figure 12. Sediment source contributions according to the sampling method predicted from the SVM-SGD-MIR models in the Conceição River catchment, Southern Brazil. Event: Rainfall runoff event; FBS: Fine bed sediment samples; TISS: Time integrated suspended sediment samples; Surface: surface sources (cropland+pasture); SB: stream bank; and UR: unpaved road.

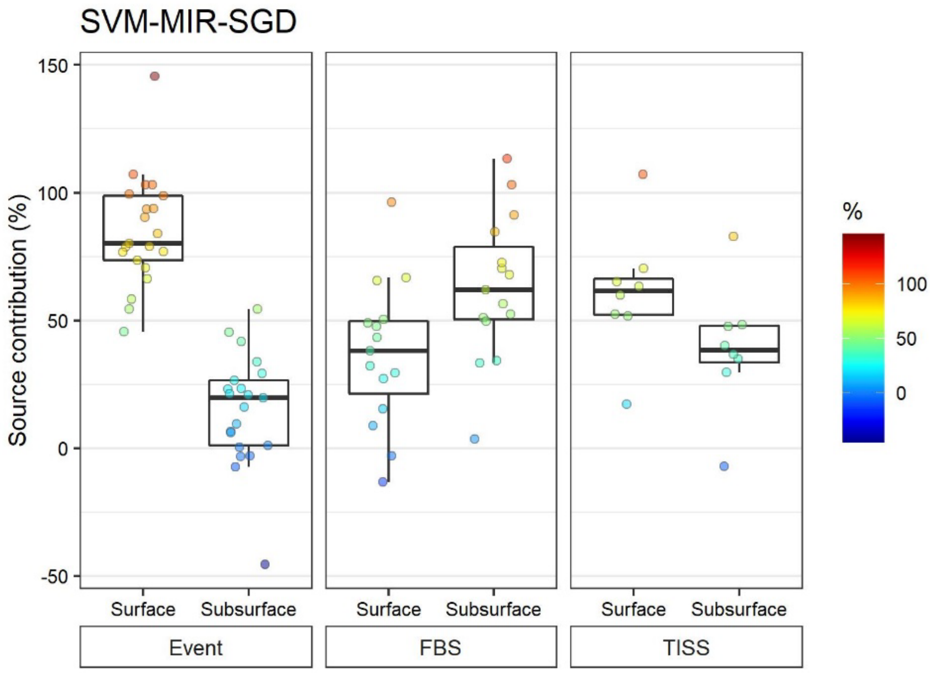

Figure 13. Sediment source contributions according to the sampling method predicted from the SVM-SGD-MIR models in the Conceição River catchment, Southern Brazil. Event: Rainfall runoff event; FBS: Fine bed sediment samples; and TISS: Time integrated suspended sediment samples.

the field and contamination of the water bodies. In contrast, the sediment collected on the riverbed shows larger contributions from the stream bank, which is in agreement with the results observed in previous studies conducted in the same catchment (Tiecher et al., 2018; Ramon et al., 2020). The reduction in source number and the combination of similar sources confirm the preliminary observations showing that surface sources provide the main suspended sediment source, for both event-based and TISS samples (Figures 12 and 13). In contrast, the subsurface sources, in general, and stream banks, in particular, provide the main source for riverbed sediment samples.

CONCLUSIONS

Diffuse reflectance spectroscopy was shown to have potential in estimating the contribution of sediment sources from a large and homogeneous subtropical catchment, and may provide a more cost-effective method than the more conventional techniques. In this environment, the support vector machine model was found to be the most powerful for estimating sediment source contributions. Ultra-violet-visible spectra were not efficient for tracing the sediment sources in this catchment with homogeneous and highly weathered soils, while near and mid-infrared showed good accuracy values. Reducing the number of potential sources reduced the uncertainties with source discrimination and model predictions. Besides the satisfactory statistical values obtained for both approaches, when a larger number of potential sources is considered, more caution with model calibration is required and testing different models, pre-processing techniques and spectral range is recommended. The approach with four sources indicates pastures provide the main source for the suspended sediment collected during events and by the time-integrated sampler. However, a high uncertainty is associated with this result. Based on the approaches with three and two sources, the model predictions indicate surface sources provide the main sediment source to suspended sediment samples collected during rainfall events and for time-integrated sediment samples, while the stream bank provides the main source to the riverbed sediment samples.

ACKNOWLEDGEMENTS

The authors are grateful to National Council for Scientific and Technological Development – CNPq and Coordination for the Improvement of Higher Education Personnel – CAPES for providing financial support. Furthermore, the authors are also grateful to CAPES for founding the PhD scholarship of the first author Rafael Ramon, in the framework of the CAPES-COFECUB Project [No. 88887.196234/2018-00]. We are also grateful to the FUNDAÇÃO DE ESTUDOS AGRÁRIOS LUIZ DE QUEIROZ – FEALQ, AGRISUS PROCESS No. 2920/20, for providing financial support. The last author also thanks the National Council for Scientific and Technological Development (CNPq) of Brazil for the research grant 311788/2019-0.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://www.rbcsjournal.org/wp-content/uploads/articles_xml/1806-9657-rbcs-48-e0230144/1806-9657-rbcs-48-e0230144-suppl01.pdf.

AUTHORS CONTRIBUTIONS

Conceptualization: Rafael Ramon (equal), Tales Tiecher (equal) and Olivier Evrard (equal).

Data curation: Rafael Ramon (lead), Tales Tiecher (equal), Olivier Evrard (equal), Jean M. Moura-Bueno (supporting), Gabriela Naibo (supporting) and Laurent Caner (supporting).

Formal analysis: Rafael Ramon (lead), Tales Tiecher (equal), Olivier Evrard (equal), Jean M. Moura-Bueno (supporting), Gabriela Naibo (supporting) and Laurent Caner (supporting).

Funding acquisition: Jean P.G. Minella (equal) and Danilo S. Rheinheimer (equal).

Resources: Tales Tiecher (equal), Olivier Evrard (equal), Jean P.G. Minella (equal), Laurent Caner (equal) and Danilo S. Rheinheimer (equal).

Writing – original draft: Rafael Ramon (lead), Tales Tiecher (supporting) and Olivier Evrard (supporting).

Writing – review editing: Rafael Ramon (equal), Tales Tiecher (equal), Olivier Evrard (equal), Jean P. G. Minella (equal), Cláudia A. P. Barros (equal), Jean M. Moura-Bueno (equal), Gabriela Naibo (equal), Laurent Caner (equal) and Danilo S. Rheinheimer (equal).

REFERENCES

Amorim F, Silva YJAB, Nascimento RC, Silva YJAB, Tiecher T, Nascimento CWA, Minella JPG, Zhang Y, Upadhayay HR, Pulley S, Collins AL. Sediment source apportionment using optical property composite signatures in a rural catchment, Brazil. Catena. 2021;202:105208. https://doi.org/10.1016/j.catena.2021.105208

Araújo SR, Wetterlind J, Demattê JAM, Stenberg B. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. Eur J Soil Sci. 2014;65:718-29. https://doi.org/10.1111/ejss.12165

Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl Spectrosc. 1989;43:772-7. https://doi.org/10.1366/0003702894202201

Bellon-Maurel V, McBratney A. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biol Biochem. 2011;43:1398-410. https://doi.org/10.1016/j.soilbio.2011.02.019

Collins AL, Blackwell M, Boeckx P, Chivers CA, Emelko M, Evrard O, Foster I, Gellis A, Gholami H, Granger S, Harris P, Horowitz AJ, Laceby JP, Martinez-Carreras N, Minella JPG, Mol L, Nosrati K, Pulley S, Silins U, Silva YJAB, Stone M, Tiecher T, Upadhayay HR, Zhang Y. Sediment source fingerprinting: benchmarking recent outputs, remaining challenges and emerging themes. J Soils Sediments. 2020;20:4160-93. https://doi.org/10.1007/s11368-020-02755-4

Collins AL, Pulley S, Foster IDL, Gellis A, Porto P, Horowitz AJ. Sediment source fingerprinting as an aid to catchment management: A review of the current state of knowledge and a methodological decision-tree for end-users. J Environ Manage. 2017;194:86-108. https://doi.org/10.1016/j.jenvman.2016.09.075

Didoné EJ, Minella JPG, Evrard O. Measuring and modelling soil erosion and sediment yields in a large cultivated catchment under no-till of Southern Brazil. Soil Till Res. 2017;174:24-33. https://doi.org/10.1016/j.still.2017.05.011

Didoné EJ, Minella JPG, Schneider FJA, Londero AL, Lefevre I, Evrard O. Quantifying the impact of no-tillage on soil redistribution in a cultivated catchment of Southern Brazil (1964–2016) with ^{137}Cs inventory measurements. Agr Ecosys Environ. 2019;284:106588. https://doi.org/10.1016/j.agee.2019.106588

Dotto AC, Dalmolin RSD, ten Caten A, Gris DJ, Ruiz LFC. AlradSpectra: A quantification tool for soil properties using spectroscopic data in R. Rev Bras Cienc Solo. 2019;43:e0180263. https://doi.org/10.1590/18069657rbcs20180263

Evrard O, Batista PVG, Company J, Dabrin A, Foucher A, Frankl A, Garcia-Comendador J, Huguet A, Lake N, Lizaga I, Martinez-Carreras N, Navratil O, Pignol C, Sellier V. Improving the design and implementation of sediment fingerprinting studies: summary and outcomes of the TRACING 2021 Scientific School. J Soils Sediments. 2022;22:1648-61. https://doi.org/10.1007/s11368-022-03203-1

Foucher A, Tassano M, Chaboche P-A, Chalar G, Cabrera M, Gonzalez J, Cabral P, Simon A-C, Agelou M, Ramon R, Tiecher T, Evrard O. Inexorable land degradation due to agriculture expansion in South American Pampa. Nat Sustain. 2023;6:662-70. https://doi.org/10.1038/s41893-023-01074-z

Golosov VN, Walling DE. Erosion and sediment problems: global hotspots. Paris, France: UNESDOC Digital Library; 2019.

IUSS Working Group WRB. World reference base for soil resources 2014, update 2015: International soil classification system for naming soils and creating legends for soil maps. Rome: Food and Agriculture Organization of the United Nations; 2015. (World Soil Resources Reports, 106).

Knox NM, Grunwald S, McDowell ML, Bruland GL, Myers DB, Harris WG. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. Geoderma. 2015;239-240:229-39. https://doi.org/10.1016/j.geoderma.2014.10.019

Laceby JP, Evrard O, Smith HG, Blake WH, Olley JM, Minella JPG, Owens PN. The challenges and opportunities of addressing particle size effects in sediment source fingerprinting: A review. Earth-Sci Rev. 2017;169:85-103. https://doi.org/10.1016/j.earscirev.2017.04.009

Legout C, Poulenard J, Nemery J, Navratil O, Grangeon T, Evrard O, Esteves M. Quantifying suspended sediment sources during runoff events in headwater catchments using spectrocolorimetry. J Soils Sediments. 2013;13:1478-92. https://doi.org/10.1007/s11368-013-0728-9

Lucà F, Conforti M, Castrignanò A, Matteucci G, Buttafuoco G. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma. 2017;288:175-83. https://doi.org/10.1016/j.geoderma.2016.11.015

Melo DCD, Anache JAA, Almeida CN, Coutinho JV, Ramos Filho GM, Rosalem LMP, Pelinson NS, Ferreira GLRA, Schwamback D, Calixto KG, Siqueira JPG, Duarte-Carvajalino JC, Jhunior HCS, Nóbrega JD, Morita AKM, Leite CMC, Guedes ACE, Coelho VHR, Wendland E. The big picture of field hydrology studies in Brazil. Hydrol Sci J. 2020;65:1262-80. https://doi.org/10.1080/02626667.2020.1747618

Mevik B-H, Wehrens R, Liland KH. Partial least squares and principal component regression. Packag. R CRAN; 2016. Available from: https://cran.r-project.org/web/packages/pls/

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, Lin C-C. Misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. Version 1.7-3. R Packag; 2019. Available from: https://cran.r-project.org/web/packages/e1071/index.html

Minella JPG, Merten GH, Barros CAP, Ramon R, Schlesner A, Clarke RT, Moro M, Dalbianco L. Long-term sediment yield from a small catchment in southern Brazil affected by land use and soil management changes. Hydrol Process. 2017;32:200-11. https://doi.org/10.1002/hyp.11404

Moura-Bueno JM, Dalmolin RSD, ten Caten A, Dotto AC, Demattê JAM. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. Geoderma. 2019;337:565-81. https://doi.org/10.1016/j.geoderma.2018.10.015

Naibo G, Ramon R, Pesini G, Moura-Bueno JM, Barros CAP, Caner L, Silva YJAB, Minella JPG, Santos DR, Tiecher T. Near-infrared spectroscopy to estimate the chemical element concentration in soils and sediments in a rural catchment. Catena. 2022;213:106145. https://doi.org/10.1016/j.catena.2022.106145

Ng W, Minasny B, Montazerolghaem M, Padarian J, Ferguson R, Bailey S, McBratney AB. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma. 2019;352:251-67. https://doi.org/10.1016/j.geoderma.2019.06.016

Ni LS, Fang NF, Shi ZH, Tan WF. Mid-infrared spectroscopy tracing of channel erosion in highly erosive catchments on the Chinese Loess Plateau. Sci Total Environ. 2019;687:309-18. https://doi.org/10.1016/j.scitotenv.2019.06.116

Poesen J. Soil erosion in the Anthropocene: Research needs. Earth Surf Process Landforms. 2017;84:64-84. https://doi.org/10.1002/esp.4250

Poulenard J, Legout C, Némery J, Bramorski J, Navratil O, Douchin A, Fanget B, Perrette Y, Evrard O, Esteves M. Tracing sediment sources during floods using Diffuse Reflectance Infrared Fourier Transform Spectrometry (DRIFTS): A case study in a highly erosive mountainous catchment (Southern French Alps). J Hydrol. 2012;414-415:452-62. https://doi.org/10.1016/j.jhydrol.2011.11.022

Poulenard J, Perrette Y, Fanget B, Quetin P, Trevisan D, Dorioz JM. Infrared spectroscopy tracing of sediment sources in a small rural watershed (French Alps). Sci Total Environ. 2009;407:2808-19. https://doi.org/10.1016/j.scitotenv.2008.12.049

Pulley S, Van der Waal B, Rowntree K, Collins AL. Colour as reliable tracer to identify the sources of historically deposited flood bench sediment in the Transkei, South Africa: A comparison with mineral magnetic tracers before and after hydrogen peroxide pre-treatment. Catena. 2018;160:242-51. https://doi.org/10.1016/j.catena.2017.09.018

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: http://www.R-project.org/.

Ramon R, Evrard O, Laceby JP, Caner L, Inda AV, Barros CAP, Minella JPG, Tiecher T. Combining spectroscopy and magnetism with geochemical tracers to improve the discrimination of sediment sources in a homogeneous subtropical catchment. Catena. 2020;195:104800. https://doi.org/10.1016/j.catena.2020.104800

Reeves III JB, Follett RF, McCarty GW, Kimble JM, Reeves JB, Follett RF, McCarty GW, Kimble JM. Can Near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? Commun. Soil Sci Plant Anal. 2006;37:2307-25. https://doi.org/10.1080/00103620600819461

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Oliveira JB, Coelho MR, Lumbreras JF, Cunha TJF. Sistema brasileiro de classificação de solos. 3. ed. rev. ampl. Rio de Janeiro: Embrapa Solos; 2013.

Sellier V, Navratil O, Laceby JP, Legout C, Foucher A, Allenbach M, Lefèvre I, Evrard O. Combining colour parameters and geochemical tracers to improve sediment source discrimination in a mining catchment (New Caledonia, South Pacific Islands). Soil. 2021;7:743-66. https://doi.org/10.5194/soil-7-743-2021

Tiecher T, Caner L, Minella JPG, Bender MA, Santos DR. Tracing sediment sources in a subtropical rural catchment of southern Brazil by using geochemical tracers and near-infrared spectroscopy. Soil Till Res. 2016;155:478-91. https://doi.org/10.1016/j.still.2015.03.001

Tiecher T, Caner L, Minella JPG, Evrard O, Mondamert L, Labanowski J, Santos DR. Tracing sediment sources using mid-infrared spectroscopy in Arvorezinha Catchment, Southern Brazil. Land Degrad Dev. 2017;28:1603-14. https://doi.org/10.1002/ldr.2690

Tiecher T, Caner L, Minella JPG, Santos DR. Combining visible-based-color parameters and geochemical tracers to improve sediment source discrimination and apportionment. Sci Total Environ. 2015;527-528:135-49. https://doi.org/10.1016/j.scitotenv.2015.04.103

Tiecher T, Minella JPG, Evrard O, Caner L, Merten GH, Capoane V, Didoné EJ, Santos DR. Fingerprinting sediment sources in a large agricultural catchment under no-tillage in Southern Brazil (Conceição River). Land Degrad Dev. 2018;29:939-51. https://doi.org/10.1002/ldr.2917

Tiecher T, Moura-Bueno JM, Caner L, Minella JPG, Evrard O, Ramon R, Naibo G, Barros CAP, Silva YJAB, Amorim FF, Rheinheimer DS. Improving the quantification of sediment source contributions using different mathematical models and spectral preprocessing techniques for individual or combined spectra of ultraviolet-visible, near- and middle-infrared spectroscopy. Geoderma. 2021;384:114815. https://doi.org/10.1016/j.geoderma.2020.114815

Viscarra Rossel RA, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma. 2010;158:46-54. https://doi.org/10.1016/j.geoderma.2009.12.025

Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma. 2006;131:59-75. https://doi.org/10.1016/j.geoderma.2005.03.007

Wijewardane NK, Ge Y, Morgan CLS. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. Geoderma. 2016;267:92-101. https://doi.org/10.1016/j.geoderma.2015.12.014